Computers in Human Behavior 27 (2011) 402–412

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management

Wade J. Mitchell, Chin-Chang Ho, Himalaya Patel, Karl F. MacDorman *

Indiana University School of Informatics, 535 West Michigan Street, IT 475, Indianapolis, IN 46202, USA

ARTICLE INFO

Keywords:
Arab–Muslim speech IAT
Audio 2AFC
Female speech IAT
IVR systems
Self-presentational bias
Social agency theory

ABSTRACT

Do people treat computers as social actors? To answer this question, researchers have measured the extent to which computers elicit social responses in people, such as impression management strategies for influencing the perceptions of others. But on this question findings in the literature conflict. To make sense of these findings, the present study proposes a dual-process model of impression management in human–computer interaction. The model predicts that, although machines may elicit nonconscious impression management strategies, they do not generally elicit conscious impression management strategies. One such strategy is presenting oneself favorably to others, which can be measured as social desirability bias when comparing self-reported preferences with implicit preferences. The current study uses both a questionnaire and an implicit association test (IAT) to compare attitudes toward human and machine speech. Although past studies on social desirability bias have demonstrated people's tendency to *underreport* their preference for the preferred group when comparing two human groups, the current study found that, when comparing human speech and machine-synthesized speech, participants instead *overreported* their preference for the preferred (human) group. This finding supports the proposed dual-process model of impression management, because participants did not consciously treat computers as social actors.

1. Introduction

Vocal interactions with technology are becoming increasingly commonplace. A customer is likely to reach an interactive voice response (IVR) system before reaching a human being when performing many everyday tasks by telephone, such as checking a bank balance, booking a flight, or contacting technical support. Voice commands are being used to control portable music players and personal digital assistants (PDAs), overcoming some of the limitations of their small screens and keyboards. Voice commands can also control in-car navigation systems, keeping the driver's visual attention on the road. With voice interfaces, users are no longer tethered to a keyboard, mouse, and display (Ghorbel et al., 2004; Martin, 1976; Möller, Krebber, & Smeele, 2006; Yerrapragada & Fisher, 1993). These kinds of hands-free interfaces can make increasingly complex technologies accessible to people with dexterity problems and other physical disabilities, including the growing population of older adults (Cohen & Oviatt, 1995; Kamm, 1994; Schafer, 1995). Voice interfaces are often features of smart houses, interactive robots, and other emerging technologies (Feil-Seifer, Skinner, & Matarić, 2007; Fong, Nourbakhsh, & Dautenhahn, 2003; Goodrich & Schultz, 2007; Helal et al., 2005; Tabar, Keshavarz, & Aghajan, 2006).

Voice interfaces can relieve us of tedious work by offering a more natural medium of control, namely, spoken language. Although the current state of the art is far from Negroponte's (1995) idealized digital butlers, IVR systems in use today continue to increase in quality, sophistication, and usefulness. Furthermore, with the broad deployment of these systems, including 4.1 billion cellular phone subscriptions worldwide (ITU, 2009), they have enormous potential for continued growth, providing additional motivation for improving IVR technology.

Efforts to improve IVR interfaces are often based on inquiries into users' experiences, such as (1) self-reports, (2) performance observation or simulation, (3) information mining of user interactions, and (4) call flow and user path design and expert evaluation (Preece, Rogers, & Sharp, 2007; Suhm, 2008; Suhm & Peterson, 2002). However, a meta-analysis of correlations among usability measures indicates that measures of user preference and satisfaction from post-test questionnaires are of little use in predicting user behavior, such as a user's efficiency and effectiveness in performing a task (Hornbæk & Law, 2007; Sauro & Lewis, 2009).[1]

* Corresponding author. Tel.: +1 317 215 7040; fax: +1 206 350 6089.
E-mail addresses: kmacdorm@indiana.edu, kfm@androidscience.com (K.F. MacDorman).

[1] Efficiency is typically operationalized by such metrics as completion time, time until event, deviation from optimal path, and frequency of use; effectiveness is typically operationalized by such metrics as error rate, binary task completion, spatial accuracy, outcome quality, completeness, and recall.

0747-5632/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2010.09.002

Other kinds of measures may be needed to improve the predictive validity of self-reports.

A problem specific to self-reports is that they may be less suited to the evaluation of IVR systems than to other kinds of interfaces. Instead of offering more typical human–computer interaction styles (e.g., windows, icons, and menus manipulated with a pointing device), a voice interface assumes the role of a human agent, encouraging people to treat it as a social actor (Nass, Steuer, & Tauber, 1994; Norman, 1994; Reeves & Nass, 1996). People come to associate a unique identity with the voice (Laurel, 1997). In this way, users anthropomorphize a system with a voice interface. Traditional usability measures provide an incomplete methodology for testing voice interfaces, because they place greater emphasis on the interface's functional aspects than on its social aspects. In other words, the measures treat the interface as a tool, not as a social actor. Because voice interfaces use spoken interaction, social aspects of spoken interaction, such as interpersonal trust, can rival more traditional usability concerns as evidenced by an increase in user acceptance of an IVR interface with a trustworthy voice (Nass & Moon, 2000; Nass, Moon, Fogg, Reeves, & Dryer, 1995).[2]

Voices for IVR systems are customarily developed using information from interviews, focus groups, and questionnaires (Kreiman et al., 2007; Tomokiyo et al., 2003). However, these sources of self-reported information often reflect the human tendency to give answers that others would view favorably. This would especially be of concern if the IVR system were being treated as a social actor. Socially desirable answers typically align with group and societal norms. This bias for social desirability is one of the symptoms of impression management, that is, the attempt to control how others perceive us (Schlenker, 1980). Social desirability bias can lead study participants to fake responses or to omit genuine responses (Ganster, Hennessey, & Luthans, 1983). Research on social desirability bias indicates that people tend to underreport their favoritism for a preferred group of people as compared to a nonpreferred group of people (Greenwald, McGhee, & Schwartz, 1998; Karpinski & Hilton, 2001; Nosek, 2005; Nosek, Greenwald, & Banaji, 2007). By activating this bias, an IVR system's perceived gender, nationality, age, or other socially sensitive characteristics could affect users' reported impressions and ratings.

Because an IVR system assumes the role of a human agent and uses a humanlike interaction style, it is natural to compare its perceived value with that of the human agent. In the evaluation of user satisfaction, social biases like in-group favoritism, mere exposure effects, and extrapersonal associations, might cause users to report a greater preference for an interface that is perceived as a human agent. However, social biases might also work against the interface if its performance were being benchmarked against that of a human being, because participants may favor fellow members of their species to a nonhuman out-group (MacDorman, Vasudevan, & Ho, 2009).

Nass et al. (1994) proposed that computers are social actors in the sense that people are inclined to treat them as if they were human (Nass & Moon, 2000; Nass et al., 1995). This basic idea has been further refined as social interface theory in human–computer interaction (Dryer, 1999) and social agency theory in computer-based education (Mayer, Sobko, & Mautone, 2003; Moreno, Mayer, Spires, & Lester, 2001). Social agency theory proposes that enhancing the social cues of an interactive agent and, in particular, the humanness of its appearance and behavior, will elicit in human users social responses found in human–human interactions (Cassell & Tartaro, 2008; Kiesler & Sproull, 1997; MacDorman & Ishiguro, 2006). Experiments testing social agency theory have found that enhancing the humanness of an agent's speech makes it not only a more likeable teacher but also a more effective one (Atkinson, Mayer, & Merrill, 2005; Mayer et al., 2003; Moreno et al., 2001). Learning outcomes improve because, according to social agency theory, social cues motivate users to make sense of the information being communicated and to apply the norms and conventions of human communication.

Despite the abovementioned research that shows that nonhuman entities often have humanlike effects on people, social desirability bias appears to function differently when machines are involved. For example, people disclose significantly more sensitive personal information on computer-administered surveys than on human-administered surveys (Tourangeau, Couper, & Steiger, 2003). Their greater willingness to share personal information indicates less social desirability bias when interacting with a computer than with a human being, because they are less motivated to present themselves favorably to the computer. These findings conflict with the predictions of the theory that computers are social actors (Nass et al., 1994). Moreover, contrary to the predictions of social agency theory, making the computer's voice sound more human did not significantly decrease self-disclosure (Couper, Singer, & Tourangeau, 2004). There are numerous other studies showing that the *conscious* knowledge that one is interacting with a computer results in a precipitous drop in impression management—for example, during an oral interview (Aharoni & Fridlund, 2007), text-based conversation (Shechtman & Horowitz, 2003), tutoring session (Bhatt, Evens, & Argamon, 2004), or bout of anger (Charlton, 2009). Until now, there has been no way to reconcile these findings with the many studies that have found that people engage in impression management when interacting with computers, for example, by displaying the same norms of politeness and reciprocity as they do with other people (Nass & Moon, 2000; Nass et al., 1994).[3]

A dual-process model of impression management in human–computer interaction may resolve these conflicting findings in the literature. When a participant answers socially sensitive questions on a survey, the relevant concepts become the focus of attention. Attentional mechanisms render the concepts conscious, and the concepts activate a sequential reasoning process in which conscious deliberation plays the primary role in suppressing observations and opinions that are contrary to social norms and conventions. By contrast, nonconscious but accessible concepts may elicit cognitive processing that is associative in nature and not semantically or logically related to the concepts (Gawronski & Bodenhausen, 2006). This kind of nonconscious processing could activate qualitatively distinct impression management strategies. Because nonconscious impression management strategies are automatic, stimulus-driven, and regulated by bottom-up processes in the brain (Frith & Frith, 2008), any computer or other mechanism producing the necessary stimulus can elicit these strategies as long as they are not consciously overridden. The proposed model holds that human beings elicit both conscious and nonconscious strategies of impression management, whereas nonhuman entities elicit only nonconscious strategies or none at all. The model is consistent with the observation that people are unaware of it when they treat computers like human beings and vehemently deny it (Nass et al., 1994; Reeves & Nass, 1996).

[2] Traditional usability concerns include compatibility, consideration for user resources (e.g., attentional limits, memorability, learnability), consistency, effectiveness, efficiency, error prevention and recovery, explicitness, feedback, prioritization of functionality and information, appropriate transfer of technology, safety, user control, utility, and visual clarity (Preece et al., 2007).

[3] The inconsistency found between computers are social actors studies and other studies on impression management in human–computer interaction may be explained in part by the fact that most computers are social actors studies did not directly compare human-computer interaction to human-human interaction. They showed that a response pattern (e.g., politeness) in human-human interaction was replicated in human-computer interaction, but they did not show the extent to which it was replicated.

The dual-process model of impression management predicts that people will exhibit social desirability bias by exaggerating their preference for human beings over machines when they apply conscious impression management strategies in explicitly comparing human speech and machine-synthesized speech. This prediction would be supported if people's favoritism for human beings over machines were found to be overstated on self-reported surveys. A baseline for such support could be provided by implicit measures. The present study is the first to use both explicit and implicit measures to investigate participants' social desirability bias in comparing human and machine-synthesized speech. The results of this study and other related studies indicate that both the theory that computers are social actors and social agency theory are in need of refinement. The dual-process model is proposed to explain conflicting findings concerning whether people engage in impression management when interacting with machines.

1.1. The purpose and approach of this study

The goal of this study is to determine whether self-reported evaluations of synthesized speech are affected by social desirability concerns and, specifically, whether a social desirability bias favoring human beings could reflect negatively on machines in self-reported evaluations. If people favored human beings over machines in their self-reported evaluations, this would imply that they were not consciously treating machines as social actors—a finding that would support the proposed dual-process model of impression management in human–computer interaction. Identifying how social desirability bias affects self-reported evaluations of voice interfaces could lead to more effective methods of developing and evaluating voices for IVR systems, enabling interaction designers to increase their user acceptance.

To achieve these goals, the implicit association test (IAT) framework of Greenwald et al. (1998) was modified to use auditory stimuli as instances of target concepts. Once validated, the auditory IAT was then used to measure the strength of implicit associations between human and machine-synthesized speech and positive and negative attributes. The association strength scores were compared with the explicit measures in this study to estimate the degree of social desirability bias for human speech vs. machine-synthesized speech.

2. Extending the implicit association test to auditory stimuli

Given the prominence of IVR systems, it is important to understand human attitudes toward machine voices. For the purposes of this study, machine voices are voices synthesized by text-to-speech software. In human–computer interaction, attitudes are typically gauged using explicit measures of user satisfaction, which rely on introspection and self-reporting. A limitation of self-reporting is that attitudes may be affected by past experiences, memories, and learned behaviors in ways the individual is unaware of or unwilling to reveal (Graf & Schacter, 1985; Greenwald, 1990; Jacoby & Dallas, 1981; Jacoby, Lindsay, & Toth, 1992; Jacoby & Witherspoon, 1982; Kihlstrom, 1990; Roediger, Weldon, & Challis, 1989; Schacter, 1987). Implicit measures estimate these associations by indirect methods, namely, through processes that are uncontrolled, unintentional, nonconscious, efficient, effortless, fast, goal-independent, autonomous, and/or stimulus-driven (De Houwer & Moors, 2007).

2.1. The implicit association test and its shortened form

Early experiments in measuring implicit associations found that response latency can be used to measure association strength (Devine, 1989; Gaertner & McLaughlin, 1983). Greenwald et al. (1998) used response latency in the original implicit association test. The IAT consists of seven sections, called blocks, but only Block 4 and Block 7 are scored; this scoring pattern is not revealed to the test's participants in advance. The scored blocks randomly interleave two 2-alternative forced choice tasks: a target concept discrimination task and an attribute dimension discrimination task. For example, one of this study's IATs uses *human* and *machine* as target concepts and *pleasant* vs. *unpleasant* as the attribute dimension.[4] During the IAT an instance of the target concepts or attributes appears on the screen in random order and must be categorized as either *human* or *machine* or as either *pleasant* or *unpleasant* by pressing either the *E* or *I* key. By measuring the strength of association of a pair of target concepts with an attribute dimension, researchers can uncover a participant's implicit associations.

In this study, a shortened version of the IAT was used to facilitate participation in multiple IATs. Greenwald, Nosek, and Banaji (2003) discovered that the 40 practice trials of Blocks 3 and 6 of the IAT had a higher correlation with the corresponding explicit measure than the 80 test trials of Blocks 4 and 7. Based on these findings, MacDorman et al. (2009) created a shortened version of the IAT by scoring Blocks 3 and 6 and eliminating Blocks 4 and 7 (Table 1). This reduced the number of blocks to five and the number of discrimination trials to 100. The shortened IAT featured an improved scoring algorithm with error latencies and IAT *D* as a measure of effect size,[5] which Greenwald et al. (2003) found resulted in higher implicit–explicit correlations than five alternative scoring algorithms. This study uses the IAT format and scoring algorithm from MacDorman et al. (2009) with some minor modifications (described in *Data analysis*). This is not the only published modification of the original IAT. Sriram and Greenwald (2009) published a different shortened version called the brief IAT.

2.2. Explicit measures

In addition to the IAT, participants complete questionnaire items about the pair of target concepts. These self-reported results are used to calculate an explicit measure, which may then be compared with the implicit measure to explore people's attitudes toward different groups of people, products, or brands (Brunel, Tietje, & Greenwald, 2004; Maison, Greenwald, & Bruin, 2004)—or, as in this study, human and machine voices. The correlation between the implicit and explicit measures is much stronger for target concepts that are not socially sensitive (e.g., *flowers* vs. *insects*) than for target concepts that are socially sensitive (e.g., *African American* vs. *white*; Dasgupta, McGhee, Greenwald, & Banaji, 2000; Greenwald et al., 1998). The IAT has been found to be resistant to social desirability bias for socially sensitive topics, such as gay and lesbian stereotypes, obesity stereotypes, gender attitudes, and attitudes toward people of particular races, nationalities, and religions (Greenwald et al., 1998; Karpinski & Hilton, 2001).

In this study, a relative preference item and the scored difference in a *warmth* index (*WD*) for each target concept served as explicitly measured counterparts to the implicit measure of the IAT. A relative preference item was included, because it forces a direct comparison between the two target concepts and, therefore, is easily affected by social desirability bias (Nosek and Greenwald, 2007). The *warmth* index was included (Ho & MacDorman, 2010), because warmth has consistently appeared in the social psychology literature as the primary dimension of interpersonal

[4] This paper focuses on the attitude IAT; however, IATs have also been devised to measure self-esteem, self-identity, and stereotypes (Greenwald et al., 2002).

[5] IAT *D* is the difference for all scored trials in the participant's mean latency in categorizing the first and the second target concept divided by the standard deviation of the latencies (Greenwald et al., 2003).

Table 1
A comparison of the original IAT and the shortened IAT.

		Original IAT		Shortened IAT	
Task	Trials	Block	Purpose	Block	Purpose
Target concept discrimination	20	1	Practice	1	Practice
Attribute discrimination	20	2	Practice	2	Practice
Combined target concept & attribute discrimination	20	3	Practice	3	Test
Combined target concept & attribute discrimination	40	4	Test		
Reverse target concept discrimination	20	5	Practice	4	Practice
Combined reverse target concept & attribute discrimination	20	6	Practice	5	Test
Combined reverse target concept & attribute discrimination	40	7	Test		

perception (Cuddy, Fiske, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002). Ensuring that both the implicit and the explicit measures are relative and affective reduces experimental design variability and thus strengthens implicit–explicit correlations (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

2.3. The auditory implicit association test

Many kinds of interpersonal judgments are based on nonverbal vocal information. In oral communication both vocal cues (e.g., pitch, intensity) and speech cues (e.g., speech rate, nonfluencies) influence listeners' interpersonal impressions (Berry, 1991). From these cues listeners are able to agree on a variety of judgments: the speaker's emotional state (Scherer & Oshinsky, 1977), personality traits (Addington, 1968; Allport & Cantril, 1934; Markel, Meisels, & Houck, 1964), race (Purnell, Idsardi, & Baugh, 1999; Walton & Orlikoff, 1994), occupation, and other attributes (Fay & Middleton, 1939). Listeners are also able to agree on which voices are attractive, and they rate speakers with attractive voices more favorably (Hughes, Harrison, & Gallup, 2002; Zuckerman, Miyake, & Hodgins, 1991). Vocal cues are more revealing of deception than facial cues and strongly influence judgments of dominance (DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Zuckerman, Amidon, Bishop, & Pomerantz, 1982). Given the impact of human vocal cues on social perception, it is important to explore their impact on perceptions of systems with machine-synthesized voices.

In unpublished work, an auditory attitude IAT was used to measure implicit associations about race using speech segments from European Americans and African Americans (Van de Kamp, 2002). In the same work, it was also used to measure attitudes, identity, and stereotypes in relation to musical genres (e.g., jazz vs. country) and ambient sounds (e.g., stock exchange vs. children's playground). Although an auditory IAT has already been used to study the lateralization of self-esteem in the brain (McKay, Arciuli, Atkinson, Bennett, & Pheils, 2010) and associations between basic tastes and pitch (Crisinel & Spence, 2009), the present study is the first *published* study to use an auditory attitude IAT and the first study to apply the IAT to machine-synthesized speech.

The IATs designed for this study use auditory stimuli to represent instances of the target concepts and visual stimuli to represent instances of the attribute dimensions. Alternating between visual and auditory stimuli during the discrimination tasks could enhance attention (Grimes, 1990; Mayer & Moreno, 1998), thus increasing the test's sensitivity.

2.4. Theoretical basis and hypotheses

The theoretical basis of the predictions made by the dual-process model of impression management in human–computer interaction entails the mediating effect of social desirability bias on in-group favoritism, extrapersonal associations, and mere exposure effects. *In-group favoritism,* which affects both attitudes and behavior, denotes the tendency to evaluate members of one's own group more favorably than nonmembers (Ashburn-Nardo, Voils, & Monteith, 2001; Mackie & Smith, 1998; Wilder & Simon, 2001). *Extrapersonal associations* denote associations between a group and an attribute that are formed by repeated exposure in the social environment, which includes the news media, educational system, popular culture, and other people's opinions (Houben & Wiers, 2007; Karpinski & Hilton, 2001; Olson & Fazio, 2004; Rudman, 2004). *Mere exposure effects* denote the influence repeated exposure to a stimulus has in increasing positive associations (Fazio & Olson, 2003; Greenwald & Banaji, 1995; Greenwald et al., 1998; Zajonc, 1968). Social desirability bias typically masks the effect of in-group favoritism, extrapersonal associations, and mere exposure effects on self-reported attitudes, which is the motivation for using implicit measures to estimate associations by means of processes that are not under conscious control (Greenwald et al., 1998; Hewstone, Rubin, & Willis, 2002; Karpinski & Hilton, 2001; Nosek & Greenwald, 2007).

In-group favoritism predicts that participants will favor their own species (Fig. 1A). Because this study's participants are human, they are predicted to have a stronger implicit association between *human speech* and *pleasant* than *machine speech* and *pleasant.* This prediction is suggested by past research showing that participants explicitly rated human speech more positively (Lee, 2010; Mayer et al., 2003; Mullennix, Stern, Wilson, & Dyson, 2003; Stern, Mullennix, Dyson, & Wilson, 1999). Mere exposure effects and, in Western popular culture,[6] extrapersonal associations also predict that participants will favor *human speech*. Consistent with the predictions of the proposed dual-process model of impression management in human–computer interaction, social desirability bias is predicted to increase favoritism for human speech over machine speech in self-reported evaluations. This prediction is supported by past research showing that participants lack social desirability bias for machine speech (Aharoni & Fridlund, 2007; Couper et al., 2004; Tourangeau et al., 2003). These predictions are operationalized in the following three hypotheses for the *human vs. machine speech* IAT:

H1A: Participants will have a stronger implicit association between *human speech* and pleasant terms than between *machine speech* and pleasant terms (because of in-group favoritism, extrapersonal associations, and mere exposure effects).
H1B: Participants will explicitly rate *human speech* more positively than *machine speech* (Lee, 2010; Mayer et al., 2003; Mullennix et al., 2003; Stern et al., 1999).
H1C: There will be no or low correlation between implicit and explicit measures comparing human speech and machine speech (because of social desirability bias).

[6] MacDorman et al. (2009) discuss how attitudes concerning robots in the West may differ from Japan in part because of their portrayal in the media and popular culture. Popular films such as *The Terminator* or *Blade Runner* may instill a negative attitude toward robots in Americans, whereas heroic robots such as Astro Boy in Japanese Manga may instill a more positive attitude towards robots in Japanese.

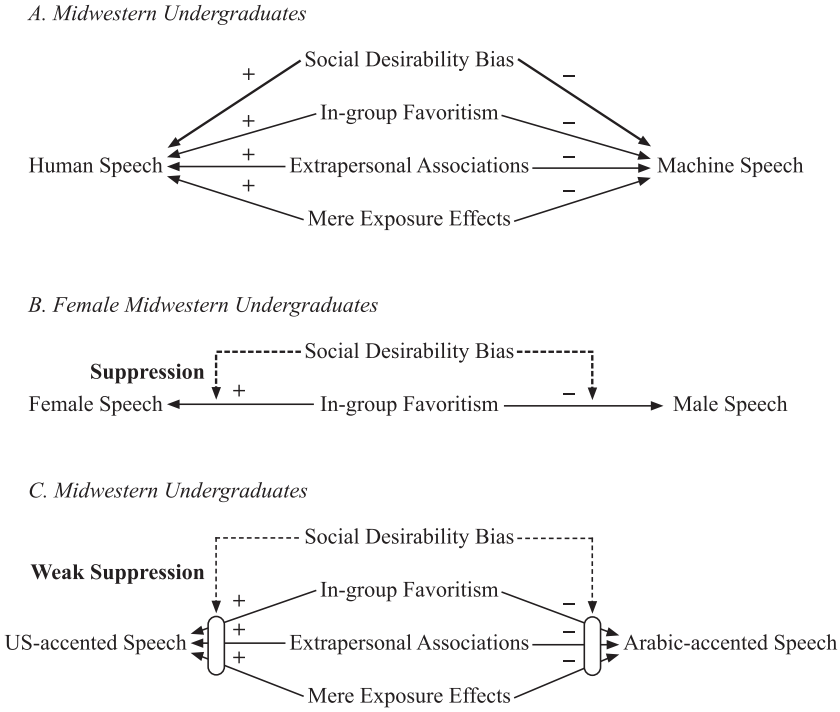

Fig. 1. When two human groups are being compared, such as females and males (B) or Americans and Arabs (C), social desirability bias typically functions to suppress the expression of positive attitudes (+) toward the preferred group and negative attitudes (−) toward the nonpreferred group. The proposed dual-process model of impression management in human–computer interaction, however, predicts that when human beings are compared with nonhuman machines (A), social desirability bias will increase the expression of positive attitudes toward the preferred group, because people do not consciously consider machines to be social actors.

To validate the shortened auditory IAT design and to provide benchmarks for interpreting the results of the *human vs. machine speech* IAT, two additional IATs were conducted on themes from the existing literature: the *female vs. male* IAT and the *US vs. Arabic-accented speech* IAT. Implicit measures indicate women strongly favor their own gender (Fig. 1B); however, men tend not to exhibit a gender preference, which is a notable exception to in-group favoritism (Nosek & Banaji, 2001; Richeson & Ambady, 2003; Rudman & Goodwin, 2004). Although studies using explicit measures have found both men and women will express a preference for women (Eagly, Mladinic, & Otto, 1994; Rudman & Goodwin, 2004), women tend to underreport their preference (Nosek & Banaji, 2001; Richeson & Ambady, 2003; Rudman & Goodwin, 2004), which is indicative of social desirability bias. The *female vs. male speech* IAT is expected to mirror these findings. These predictions are operationalized in the following four hypotheses:

H2A: For female participants, the implicit association between *female speech* and pleasant terms will be stronger than between *male speech* and pleasant terms.
H2B: For female participants, there will be no or low correlation between implicit and explicit measures of gender bias (Greenwald & Farnham, 2000).
H2C: For male participants, the implicit association between *female speech* and pleasant terms will not be significantly stronger or weaker than between *male speech* and pleasant terms.
H2D: For male participants, there will be no or low correlation between implicit and explicit measures of gender bias (Greenwald & Farnham, 2000).

The *US vs. Arabic-accented speech* IAT was selected to validate the shortened IAT, because the literature indicates reduced social desirability bias among Americans while rating Arabic concepts (Oswald, 2005). In other words, Americans are relatively willing to state explicitly negative associations with Arabic people. Coupling this known bias and a reduced moderating effect of social desirability makes Arabic target concepts an effective means of gauging the validity of the IAT designs proposed in this study based on the correlation between implicit and explicit measures.

In-group favoritism predicts that Midwestern undergraduates would favor US-accented speech over Arabic-accented speech (Fig. 1C). In-group favoritism tends to be especially strong if the other group is perceived as a threat (Oswald, 2005). Extrapersonal associations predict that anti-Arab sentiments could be formed by exposure to news that focused on Arab Muslim extremists in the context of terrorism, especially in the aftermath of the September 11, 2001 attacks on the World Trade Center in New York City (Park, Felix, & Lee, 2007). The mere exposure effect predicts that Midwestern undergraduates will have stronger implicit associations between *US* and *pleasant* than *Arabic* and *pleasant,* because US accents are more familiar to them than Arabic ones.[7] These predictions are operationalized in the following two hypotheses:

H3A: Participants will have a stronger implicit association between *US-accented speech* and pleasant terms than between *Arabic-accented speech* and pleasant terms.
H3B: There will be a medium-to-high correlation between implicit and explicit measures comparing *US-accented speech* and *Arabic-accented speech*.

[7] As US society is not segregated by gender, it is less clear how mere exposure affects the *female vs. male speech* IAT, though maternal bonding from early infancy might create a preference for women (Rudman & Goodwin, 2004).

3. Methods

3.1. Participants

Participants were recruited in May 2008 from a list of randomly selected undergraduate students and recent graduates of a nine-campus Midwestern university. Additional inclusion criteria were age 18 or older, born in the United States, and current US residency. Among a total of 485 participants, 151 (31%) were male, 334 were female (69%), 100 (21%) were under 20 years old, 218 (45%) were 21 to 25 years old, and 167 (34%) were over 26 years old. In this between-groups design, there were 167 participants in the *human vs. machine speech* IAT, 180 participants in the *female vs. male speech* IAT, and 138 participants in the *US vs. Arabic-accented speech* IAT. To characterize the sample's representativeness of the undergraduate population as a whole (excluding foreigners) for the three IATs, the measurement error range was ±7.6%, ±7.3%, and ±8.3%, respectively, at a 95% confidence level. Except for the absence of foreign participants, the participants reflected the demographics of the university's undergraduate population (80.1% non-Hispanic white, 6.9% African-American, 3.4% Asian, 3.0% Hispanic, and 6.6% foreign or unclassified).

3.2. Materials

The *human vs. machine speech* IAT used the concept labels *human* and *machine*; the *female vs. male speech* IAT used the concept labels *female* and *male*; and the Arabic vs. US-accented speech IAT used the concept labels *Midwestern* and *Arab*. To help convey instances of these three concept labels pairs in the IATs, the following 16 neutral terms were used: *candle holder*, *cardboard box*, *ceiling fan*, *coffee cup*, *glass bottle*, *ironing board*, *living room*, *magazine rack*, *mixing bowl*, *peanut butter*, *piano bench*, *picket fence*, *plastic cup*, *remote control*, *television set*, and *vacuum cleaner.*

The audio recordings of the three IATs differed according to whether the speaker was human or computer, female or male, and Midwestern or Arab, respectively. The *human vs. machine speech* IAT and the *male vs. female speech* IAT presented the neutral terms in eight distinct voices: two male and two female Midwestern US-accented native English speakers and two male and two female machine-synthesized voices. The male machine voices were represented by Microsoft Mike and the ReadPlease male voice, and the female machine voices were represented by Microsoft Mary and the ReadPlease female voice.[8] The *US vs. Arabic-accented speech* IAT presented the neutral terms in four distinct voices: Two were of male Midwestern US-accented native English speakers, and two were of male Arabic-accented English-as-a-second-language speakers.

All three auditory IATs used the same attribute dimension labels: *pleasant* and *unpleasant.* They also used the same set of 24 words to represent instances of the attribute dimensions. The 12 positive instances of the attribute dimension were *wonderful*, *glorious*, *happy*, *love*, *good*, *pleasure*, *success*, *peace*, *joy*, *laughter*, *affection*, and *ecstasy*, and the 12 negative instances were *horrible*, *shameful*, *sad*, *hate*, *evil*, *pain*, *failure*, *nasty*, *awful*, *hurt*, *war*, and *agony*. These 24 words were rated as either high or low in pleasantness by college students (Bellezza, Greenwald, & Banaji, 1986) and are commonly used in attitude IATs.

The questionnaire for the explicit measures was composed of the relative preference item and the *warmth* index. The relative preference item ranged from +3 ("I strongly prefer *target concept A* to *target concept B*") to −3 ("I strongly prefer *target concept B* to *target concept A*"). The *warmth* index was composed of nine 7-point semantic differential items, which ranged from −3 to +3 with 0 as neutral: *warm–cold, friendly–hostile, well-intentioned–spiteful, good-natured–grumpy, sincere–phony, happy–sad, good–bad, wonderful–terrible,* and *pleased–annoyed.*

3.3. Procedures

The IAT experiments were conducted at a website.[9] Each participant registered at the website, provided demographic data, and gave informed consent. An instance of a target concept was presented as audio, and an instant of an attribute dimension was presented as text in the middle of the browser window. Concept labels and attribute dimension labels appeared in the upper left and upper right areas of the browser window.

Each IAT was followed by a questionnaire comprised of a relative preference item and the semantic differential scales of the *warmth* index. The order of the items in the semantic differential scales was random for each participant. The presentation order of the attribute–concept pairings within each IAT was counterbalanced.

3.4. Data analysis

The IAT *D* score was calculated using the scoring algorithm from Greenwald et al. (2003) with some exceptions. First, five blocks were used instead of seven (Table 1; MacDorman et al., 2009). Second, the Greenwald et al. (2003) scoring algorithm truncated results so that any response latency below 300 ms was recoded as 300 ms and any response latency above 3000 ms was recoded as 3000 ms. The present study eliminated the extreme responses rather than truncating them. Furthermore, a participant's response set was completely removed if more than 10% of the trial latencies fell outside a 300–3000 ms range. Third, the present study removed the first trial of Blocks 3 and 5, because response latencies for these trials often exceeded 3000 ms (40.6–66.5% in Block 3 and 21.5–47.5% in Block 5).

Relative preference is the mean of participants' responses to the relative preference item. The Warmth Difference was calculated by averaging the results from each semantic differential scale with respect to the two target concepts and taking the difference. Internal reliability and correlation analysis were performed using SPSS. Correlations were calculated using Pearson's correlation coefficient, and statistical significance was calculated using a two-tailed *t*-test with an alpha level of .05.

4. Results

Table 2 presents the reliability of the IAT *D* and *warmth* index for the three IATs. Cronbach's αs exceeded the standard .7 threshold, demonstrating high reliability, except for IAT *D* in the *US vs. Arabic-accented speech* IAT.

Table 3 presents the mean and standard deviation of IAT *D*, IAT Cohen's *d*, the mean and standard deviation of relative preference item, and the mean, standard deviation, Cohen's *d*, and *r*-squared of the Warmth Difference. A one-sample *t*-test confirmed that all IAT *D* scores were highly significant ($p \leqslant .001$). The positive IAT *D* scores and IAT Cohen's *d* indicate that, as predicted from the literature, across all IATs participants had stronger implicit associations between the first target concept (e.g., *human speech*) and *pleasant* than between the second target concept (e.g., *machine speech*) and *pleasant*. These results indicate the success of the shortened IAT.

[8] For all IATs, there were a sufficient number of instances of each target concept not to impact the results (Nosek, Greenwald, & Banaji, 2005).

[9] http://experiment.informatics.iupui.edu.

Table 2
Reliability of IAT *D* and the warmth index.

Topic	IAT *D*	Warmth index
	Cronbach's α	Cronbach's α
Human vs. machine speech	.78	.95
Female vs. male speech	.78	.93
US vs. Arabic-accented speech	.60	.93

Evidence from the literature indicates a social desirability bias favoring *human speech* in self-reported evaluations (Aharoni & Fridlund, 2007; Couper et al., 2004; Tourangeau et al., 2003). The explicit results are much more in favor of *human speech* than the implicit results would predict. If a weak correlation between implicit and explicit results can be attributed to social desirability bias as the literature suggests (Nosek, 2005), the results of the *human vs. machine speech* IAT indicate a strong social desirability bias in favor of *human speech.*

The IAT *D* score ($M = 0.33$, $SD = 0.30$) indicates a stronger implicit association between *human speech* and *pleasant* than *machine speech,* but compared with the other IATs in this study, the IAT *D* score is not the largest. The *RP* item indicated a very strong preference for *human speech* ($M = 2.37$, $SD = 0.86$) as did the results from the *WD* ($d = 2.30$, $r^2 = .76$). *Human speech* received the highest *warmth* index rating among all three IATs ($M = 1.25$, $SD = 0.88$), and *machine speech* received the lowest ($M = -0.65$, $SD = 0.76$). Of the target concepts, only *machine speech* received a negative *warmth* rating. There is no significant correlation between the IAT *D* score and the *RP* item ($r = .09$, $p = .234$) and only a weak correlation between the IAT *D* score and the *WD* ($r = .16$, $p = .037$; Table 4). The strong explicit results compared with the average IAT *D* score indicate a strong social desirability bias favoring *human speech.*

Previous research by MacDorman et al. (2009) using a visual IAT found similar results. Implicit results in their study showed a stronger association between *humans* and *pleasant* than *robots* for US participants ($M = 0.40$). Explicit results of their relative preference item ($M = 2.23$) and *warmth* scale difference ($M = 0.94$) are also comparable with the *human vs. machine speech* IAT of this study.

H1A and H1B predicted that participants would favor *human speech* over *machine speech* on implicit and explicit measures (Lee, 2010; Mayer et al., 2003; Mullennix et al., 2003; Stern et al., 1999). However, H1C predicted that there would be no or low correlation between implicit and explicit measures. The results of the *human vs. machine speech* IAT support these hypotheses. The IAT *D* score shows a stronger implicit association between *human speech* and *pleasant* than *machine speech* ($M = 0.33$, $SD = 0.30$). Possible explanations of this social bias include in-group favoritism (i.e., *speciesism*), extrapersonal associations (e.g., acquired from negative experiences with IVR systems or negative depictions of machines in the media; MacDorman et al., 2009), and mere exposure effects caused by the ubiquity of human speech. However, the implicit association strength between *human speech* and *pleasant* for all participants was less than the implicit association strength between *female speech* and *pleasant* for female participants in the *female vs. male speech* IAT ($M = 0.42$, $SD = 0.36$). Nevertheless, the explicit *RP* and *WD* measures indicated a very strong preference for *human speech*; its *RP* measure across all participants was 22 times greater than the *RP* measure for female participants in the *female vs. male speech* IAT (2.37 vs. 0.14; Table 3). However, the *RP* measure was not significantly correlated with IAT *D*, and *WD* was only weakly correlated. In sum, the results of the IATs indicate participants tend to underreport their favoritism for a preferred group of people as compared with a nonpreferred group of people but they tend to overreport their favoritism for human beings as compared with nonhuman machines.

The results of the *female vs. male speech* IAT demonstrate that the proposed auditory IAT, combined with the explicit measure in this study, can be used effectively to estimate social desirability bias consistent with findings in the literature. The literature shows that women have more positive implicit associations with women than with men, whereas men are neutral (Nosek & Banaji, 2001; Rudman & Goodwin, 2004). It also shows that both women and men express preference for women on explicit measures (Eagly et al., 1994). The *female vs. male speech* IAT was the only IAT in this study with a significant gender difference, so results from male and female participants were analyzed separately.

The implicit and explicit measures indicated a social desirability bias among both female and male participants consistent with the

Table 3
Mean and standard deviation of IAT *D*, IAT Cohen's *d*, mean and standard deviation of relative preference, and mean, standard deviation, Cohen's *d*, and *r*-squared of the Warmth Difference.

Topic	IAT			RP		WD			
	M	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	r^2
Human vs. machine speech ($n = 167$)	0.33	0.30	0.30	2.37	0.86	1.90	1.20	2.30	.76
Female vs. male speech, males ($n = 51$)	0.02	0.31	0.02	0.82	1.52	0.9	1.16	0.83	.38
Female vs. male speech, females ($n = 129$)	0.42	0.36	0.46	0.14	1.43	0.27	1.04	0.31	.15
US vs. arabic-accented speech ($n = 138$)	0.18	0.36	0.20	0.77	1.48	0.15	1.17	0.17	.08

Table 4
IAT *D* to relative preference correlation, IAT *D* to Warmth Difference correlation, relative preference to IAT *D* ratio, and Warmth Difference Cohen's *d* to IAT *D* ratio.

Topic	*D* to *RP*	*D* to *WD*	*RP* to *D*	*d* to *D* Ratio
	Corr.	Corr.	Ratio	
	($r_{D,RP}$)	($r_{D,WD}$)	(*RP*/*D*)	(*d*/*D*)
Human vs. machine speech ($n = 167$)	.09	.16*	7.29	7.08
Female vs. male speech, males ($n = 51$)	.20	.10	42.67	41.32
Female vs. male speech, females ($n = 129$)	.20*	.16	0.33	0.74
US vs. arabic-accented speech ($n = 138$)	.34***	.36***	4.27	0.93

* $p < .05$.
*** $p < .001$ (two-tailed).

literature. The IAT *D* score shows that female participants had a stronger implicit association between *female speech* and *pleasant* than *male speech* ($M = 0.43$, $SD = .36$), whereas male participants were neutral ($M = .02$, $SD = .31$, $F(1178) = 49.12$, $p = .000$). Moreover, female participants often underreported their in-group favoritism on the explicit measure *RP* ($M = 0.14$, $SD = 1.43$) and *WD* ($M = 0.27$, $SD = 1.04$), whereas male participants reported an even stronger preference for woman than female participants did on the explicit measure *RP* ($M = 0.82$, $SD = 1.52$, $F(1178) = 8.03$, $p = .005$) and *WD* ($M = 0.90$, $SD = 1.16$, $F(1178) = 12.57$, $p = .001$), although male participants were neutral on the implicit measures. Female participants may have felt embarrassed by their in-group favoritism, whereas male participants may have felt that reporting warmth for other men could be misinterpreted as their sexual preference.

H2A predicted that female participants would have a stronger implicit association between *female speech* and *pleasant* than *male speech,* though H2B predicted that for female participants, there would be no or low correlation between implicit and explicit measures of gender bias (Greenwald & Farnham, 2000). The results for female participants from the *female vs. male speech* IAT support these hypotheses. H2C predicted that male participants would not have a significantly stronger or weaker association between *female speech* and *pleasant* than *male speech.* In addition, H2D predicted that for male participants, there would be no or low correlation between implicit and explicit measures of gender bias (Greenwald & Farnham, 2000). The results of the male participants from the *female vs. male speech* IAT support these hypotheses.

The results of this study show positive associations by female participants (IAT Cohen's $d = 0.46$) and indifference by male participants (IAT Cohen's $d = 0.02$). Nosek and Banaji (2001) showed similar results in their gender IAT using the Go/No-go Association Test (GNAT). Female participants had more positive implicit associations with women than with men (IAT Cohen's $d = 1.02$), whereas male participants were relatively neutral (IAT Cohen's $d = 0.21$). Richeson and Ambady (2003) also conducted an attitude IAT concerning gender associations with good and bad attributes and reported results, given as the mean log of response latency, that were comparable with those of the *female vs. male speech* IAT of this study.

The correlations in the *female vs. male speech* IAT (Table 4) agree with results from the literature. Male participants had no significant correlation between the IAT *D* score and the *RP* item ($r = .20$, $p = .166$) or the IAT *D* score and the *WD* ($r = .10$, $p = .502$). Female participants had a significant but weak correlation between the IAT *D* score and the *RP* measure ($r = .20$, $p = .025$) and no significant correlation between the IAT *D* score and the *WD* ($r = .16$, $p = .063$). Nosek and Banaji (2002) also found no significant correlation between the implicit results of the GNAT and an explicit measure of preference ($r = .13$, $p > .05$).

In the *US vs. Arabic-accented speech* IAT, the IAT *D* score indicated a stronger association between *US accents* and *pleasant* than *Arabic accents*. Table 4 reports the correlations and ratios between the IAT *D* scores and the means of the explicit measure. The correlation between the IAT *D* score and the *RP* item was significant ($p = .000$) and the effect size was medium ($r = .34$) as was the correlation between the IAT *D* score and the *WD* ($p = .000$, $r = .36$). The significant correlation between the IAT *D* score and the *RP* in the *US vs. Arabic-accented speech* IAT provides evidence that auditory features can be used as instances of the target concept pair in the discrimination task. H3A predicted that participants would have a significantly stronger implicit *pleasant* association with *US-accented speech* than with *Arabic-accented speech*. The results of the *US vs. Arabic-accented speech* IAT support this hypothesis. H3B predicted a medium-to-high correlation between implicit and explicit measures, which was also supported.

5. General discussion

The current study compared the implicit measures of the IAT with explicit measures calculated from self-reported data. A usual pattern of response indicating social desirability bias favoring the nonpreferred group did *not* appear in the *human vs. machine speech* IAT. In fact, the results supported precisely the opposite claim: the large explicit-to-implicit ratio indicates participants were exaggerating their preference for human beings (Table 4). Thus, contrary to the results of other intergroup attitude IATs, the reporting bias favored the group that earned more positive implicit associations. The correlation between implicit and explicit measures ($r_{D,RP} = .09$) was much lower than the average for IATs comparing different groups of people ($\rho = .253$) or different products ($\rho = .336$) based on a meta-analysis (Hofmann et al., 2005).

The results of the *human vs. machine speech* IAT support the predictions of the proposed dual-process model of impression management in human–computer interaction. They indicate that the conscious impression management strategies activated by the explicit measures favored the preferred group, human beings, and not machines. This finding disputes the trend found in social psychology studies comparing two human groups, in which social desirability bias enhanced ratings of the nonpreferred group. The results are consistent with a number of studies that have found participants display little or no social desirability bias toward computers (Aharoni & Fridlund, 2007; Bhatt et al., 2004; Charlton, 2009; Couper et al., 2004; Shechtman & Horowitz, 2003; Tourangeau et al., 2003). They suggest that the theory that computers are social actors and social agency theory are only supported by behavioral data elicited by nonconscious impression management strategies.

An example of one such strategy is the elicitation of prosocial behavior from the presence of subtle cues of being observed in the environment, such as the mere inclusion of eyes in an image on a computer screen (Burnham & Hare, 2007; Haley & Fessler, 2005). In a study conducted in an unattended university coffee room, payment compliance with a drinks pricelist almost tripled if the pricelist included a photograph of human eyes (Bateson, Nettle, & Roberts, 2006). The cognitive processes underlying this kind of prosocial behavior are nonconscious, automatic, and stimulus-driven: Participants are completely unaware of the relation between the depiction of eyes and the change in their behavior. In another example of nonconscious impression management, participants were unaware that they rated the performance of a tutoring program more favorably if they rated it on the same computer on which they had been tutored (Nass et al., 1994). If the computer had been a human tutor, this would be interpreted as typical politeness behavior. The disparity in the results of studies that appear to support or refute social desirability bias for machines could be explained if people predominantly manage their impressions with machines by means of nonconscious strategies while refraining from using conscious strategies. The proposed dual-process model can explain most of the data of proponents of the theory that computers are social actors, such as the fact that, when people are being polite to computers, they are not even aware of it (Nass & Moon, 2000; Nass et al., 1994).

5.1. Limitations and future work

The main limitation of this study is that variables other than social desirability bias can act as moderators between implicit and explicit measures. These variables include evaluative strength, dimensionality, distinctiveness (Nosek, 2005), additional information integration for explicit representations, and research design factors (Hofmann et al., 2005). Although care was taken in the de-

sign of the implicit and explicit measures and choice of the benchmarking topics to reduce the effects of these variables, further research is required to determine the extent of their influence and to control for it. To achieve this, future experiments will include other behavioral measures that enable the assessment of predictive validity (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), including additional measures of social desirability and absolute measures of implicit preference, such as the single-target IAT (Bluemke & Friese, 2008).

A possible limitation with using the current design in developing machine-synthesized voices concerns the extent to which the IAT *D* score reflects positive or negative associations with the auditory stimuli as opposed to the concept labels under which the stimuli are categorized. Although the IAT *D* score was intended to measure the relative valence of a pair of target concepts, several studies have established that it is also affected by stimuli valence (Bluemke & Friese, 2006; Govan & Williams, 2004), though the extent of the influence has been disputed (De Houwer, 2001). To evaluate the valence of a particular machine-synthesized voice, it would be best to use an IAT or other implicit measure that does not use target concept labels during the categorization task. This would increase the salience of the stimuli and focus attention on them. The single-target IAT may easily be adapted to this purpose.

6. Conclusion

This study examined whether impression management influences self-reported evaluations of machine-synthesized speech. In the *human vs. machine speech* IAT, explicit measures derived from self-reported evaluations were benchmarked against implicit measures, and the explicit-to-implicit ratio and correlation were compared with those of IATs with human groups. The results indicate impression management can create strong favoritism for human beings over machines in explicit measures. Participants were inclined to overreport their favoritism for the preferred human group rather than underreport it, as is typical when two human groups are being compared.

The fact that impression management was directed at human beings but not computers does not support the theory that computers are social actors (Nass et al., 1994), social interface theory (Dryer, 1999), or social agency theory (Moreno et al., 2001). However, the proposed dual-process model of impression management can explain the discrepancy between this study's findings and those of the proponents of these theories, because it stipulates that computers and other nonhuman entities can more easily elicit nonconscious impression management strategies than conscious ones.

When participants exhibit social desirability bias for or against a machine-synthesized voice, the information they provide is unlikely to reflect their attitudes accurately. The use of implicit measures to detect social desirability bias could provide designers with information that more completely explains user preferences and better predicts user behavior. With this information designers can create machine voices that are more appealing to their users. Such voices would be beneficial in many areas of human–machine interaction, from IVR systems to socially assistive robots.

Acknowledgments

The authors would like to express their gratitude to the anonymous reviewers and the following researchers (mainly social psychologists) for their thoughtful suggestions on improving earlier drafts of this paper: Eyal Aharoni, Leslie Ashburn-Nardo, Anne-Sylvie Crisinel, Stephen J. Cowley, Alexander Y. Fedorikhin, Greg Francis, Anthony Greenwald, Jan De Houwer, Eun-Ju Lee, Nicole Shechtman, Jeff Sherman, Reinout W. Wiers, and Kipling D. Williams. We would also like to thank Sandosh Vasudevan for his assistance in creating the IAT website for this study. This study has been approved by the IUPUI/Clarian Research Compliance Administration (EX0805-11B) and was supported by an IUPUI Signature Center grant.

References

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35*(4), 492–503.
Aharoni, E., & Fridlund, A. J. (2007). Social reactions toward people vs. computers: How mere labels shape interactions. *Computers in Human Behavior, 23*(3), 2175–2189.
Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *Journal of Social Psychology, 5*, 37–55.
Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*(5), 789–799.
Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology, 30*(1), 117–139.
Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biological Letters, 2*(3), 412–414.
Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments and Computers, 18*(3), 299–303.
Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*(2), 298–307.
Bhatt, K., Evens, M., & Argamon, S. (2004). Hedged responses and expressions of affect in human/human and human/computer tutorial interactions. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. August 5–7, Chicago.
Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology, 42*(2), 163–176.
Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology, 38*(6), 977–997.
Brunel, F. F., Tietje, B. C., & Greenwald, A. G. (2004). Is the implicit association test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology, 14*(4), 385–404.
Burnham, T. C., & Hare, B. (2007). Engineering human cooperation. *Human Nature, 18*(2), 88–108.
Cassell, J., & Tartaro, A. (2008). Intersubjectivity in human–agent interaction. *Interaction Studies, 8*(3), 391–410.
Charlton, J. (2009). The determinants and expression of computer-related anger. *Computers in Human Behavior, 25*(6), 1213–1221.
Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human–machine communication. *Proceedings of the National Academy of Sciences, 92*(22), 9921–9927.
Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics, 20*(3), 551–570.
Crisinel, A.-S., & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience Letters, 464*(1), 39–42.
Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631–648.
Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*(3), 316–328.
De Houwer, J. (2001). A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology, 37*(6), 443–451.
De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 179–194). New York: Guilford Press.
DePaulo, B. M., Rosenthal, R., Eisenstat, R. A., Rogers, P. L., & Finkelstein, S. (1978). Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology, 36*(3), 313–323.
Devine, P. G. (1989). Stereotypes and prejudice. Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.
Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence, 13*(2), 273–295.
Eagly, A. H., Mladinic, A., & Otto, S. (1994). Cognitive and affective bases of attitudes toward social groups and social policies. *Journal of Experimental Social Psychology, 30*(2), 113–137.
Fay, P. J., & Middleton, W. C. (1939). Judgment of occupation from the voice as transmitted over a public address system and over a radio. *Journal of Applied Psychology, 23*(5), 586–601.
Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297–327.
Feil-Seifer, D., Skinner, K., & Matarić, M. (2007). Benchmarks for evaluating socially assistive robotics. *Interaction Studies, 8*(3), 423–439.
Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from

perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.
Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3–4), 143–166.
Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron, 60*(3), 503–510.
Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46*(1), 23–30.
Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *The Academy of Management Journal, 26*(2), 321–331.
Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692–731.
Ghorbel, M., Segarra, M.-T., Kerdreux, J., Keryell, R., Thepaut, A., & Mokhtari, M. (2004). Networking and communication in smart home for people with disabilities. In K. Miesenberger, J. Klaus, W. Zagler, & D. Burger (Eds.), *ICCHP 2004: Proceedings of the Ninth International Conference on Computers Helping People* (pp. 937–944). Berlin: Springer-Verlag.
Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. *Foundations and Trends in Human–Computer Interaction, 1*(3), 203–275.
Govan, C. L., & Williams, K. D. (2004). Reversing or eliminating IAT effects by changing the affective valence of the stimulus items. *Journal of Experimental Social Psychology, 40*(4), 357–365.
Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*(3), 501–518.
Greenwald, A. G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society, 28*(3), 254–260.
Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.
Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3–25.
Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022–1038.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.
Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216.
Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.
Grimes, T. (1990). Audio–video correspondence and its role in attention and memory. *Educational Technology Research and Development, 38*(3), 15–25.
Haley, K. J., & Fessler, D. M. (2005). Nobody's watching: Subtle cues affect generosity in an anonymous economic game? *Evolution and Human Behavior, 26*(3), 245–256.
Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y., & Jansen, E. (2005). The Gator Tech Smart House: A programmable pervasive space. *Computer, 38*(3), 50–60.
Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*(1), 575–604.
Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*(6), 1508–1518.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369–1385.
Hornbæk, K., & Law, E. L.-C. (2007). Meta-analysis of correlations among usability measures. In *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems* (pp. 617–626). New York: ACM.
Houben, K., & Wiers, R. W. (2007). Are drinkers implicitly positive about drinking alcohol? Personalizing the alcohol-IAT to reduce negative extrapersonal contamination. *Alcohol and Alcoholism, 42*(4), 301–307.
Hughes, S. M., Harrison, M. A., & Gallup, G. G. Jr., (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior, 23*(3), 173–180.
ITU (2009). *Measuring the information society: The ICT development index*. Geneva: International Telecommunications Union.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*(3), 306–340.
Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist, 47*(6), 802–809.
Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36*(2), 300–324.
Kamm, C. (1994). User interfaces for voice applications. In D. B. Roe & J. G. Wilpon (Eds.), *Voice communication between humans and machines* (pp. 422–442). Washington, DC: National Academy Press.
Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*(5), 774–788.
Kiesler, S., & Sproull, L. (1997). "Social" human–computer interaction. In B. Friedman (Ed.), *Human values and the design of technology* (pp. 191–199). Stanford, CA: CLSI Publications.
Kihlstrom, J. F. (1990). The psychological unconscious. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (1st ed., pp. 424–442). New York: Guilford Press.
Kreiman, J., Gerratt, B. R., & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *The Journal of the Acoustical Society of America, 122*(4), 2354–2364.
Laurel, B. (1997). Interface agents: Metaphors with character. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 207–219). Stanford, CA: CSLI Publications.
Lee, E.-J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior, 26*(4), 665–672.
MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies, 7*(3), 297–337.
MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23*(4), 485–510.
Mackie, D. M., & Smith, E. R. (1998). Intergroup relations: Insights from a theoretically integrative approach. *Psychological Review, 105*(3), 499–529.
Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology, 14*(4), 405–415.
Markel, N. N., Meisels, M., & Houck, J. E. (1964). Judging personality from voice quality. *Journal of Abnormal and Social Psychology, 69*(4), 458–463.
Martin, T. B. (1976). Practical applications of voice input to machines. *Proceedings of the IEEE, 64*(4), 487–501.
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.
Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*(2), 419–425.
McKay, R., Arciuli, J., Atkinson, A., Bennett, E., & Pheils, E. (2010). Lateralisation of self-esteem: An investigation using a dichotically presented auditory adaptation of the implicit association test. *Cortex, 46*(3), 367–373.
Möller, S., Krebber, J., & Smeele, P. (2006). Evaluating the speech output component of a smart-home system. *Speech Communication, 48*(1), 1–27.
Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177–213.
Mullennix, J. W., Stern, S. E., Wilson, S. J., & Dyson, C. (2003). Social perception of male and female computer synthesized speech. *Computers in Human Behavior, 19*(4), 407–424.
Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the AMC/SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). New York: ACM.
Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human–Computer Studies, 43*(2), 223–239.
Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.
Negroponte, N. (1995). *Being digital*. New York: Knopf.
Norman, D. A. (1994). How might people interact with agents? *Communications of the ACM, 37*(7), 68–71.
Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*(4), 565–584.
Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625–666.
Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166–180.
Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292). New York: Psychology Press.
Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653–667.
Oswald, D. L. (2005). Understanding anti-Arab reactions post-9/11: The role of threats, social categories, and personal ideologies. *Journal of Applied Social Psychology, 35*(9), 1775–1799.
Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes toward Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology, 29*(1), 35–45.
Preece, J., Rogers, Y., & Sharp, H. (2007). *Interaction design: Beyond human–computer interaction* (2nd ed.). New York: Wiley.
Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*(1), 10–30.
Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology, 39*, 177–183.

Roediger, H. L., III., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III. & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–42). Hillsdale, NJ: Erlbaum.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Assessing sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: Johns Hopkins University Press.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79–82.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personal and Social Psychology, 87*(4), 494–509.

Sauro, J., & Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. In *Proceedings of the 27th International ACM/SIGCHI Conference on Human Factors in Computing Systems* (pp. 1609–1618). Boston, MA, New York: ACM.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*(3), 501–518.

Schafer, R. W. (1995). Scientific bases of human–machine communication by voice. *Proceedings of the National Academy of Sciences, 92*(22), 9914–9920.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331–346.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations*. Belmont, Calif.: Brooks/Cole.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 281–288). Fort Lauderdale, Florida.

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56*(4), 283–294.

Stern, S. E., Mullennix, J. W., Dyson, C., & Wilson, S. J. (1999). The persuasiveness of synthetic speech versus human speech. *Human Factors, 41*(4), 588–595.

Suhm, B. (2008). IVR usability engineering using guidelines and analyses of end-to-end calls. In D. Gardener-Bonneau & H. E. Blanchard (Eds.), *Human factors and voice interactive systems* (2nd ed., pp. 1–41). New York: Springer Science.

Suhm, B., & Peterson, P. (2002). A data-driven methodology for evaluating and optimizing call center IVRs. *International Journal of Speech Technology, 5*(1), 23–37.

Tabar, A. M., Keshavarz, A., & Aghajan, H. (2006). Smart home care network using sensor fusion and distributed vision-based reasoning. In *VSSN '06: Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks* (pp. 145–154). New York: ACM.

Tomokiyo, L. M., Black, A. W., & Lenzo, K. A. (2003). Arabic in my hand: Small-footprint synthesis of Egyptian Arabic. In *Proceedings of Eurospeech 2003* (pp. 2049–2052). Geneva, Switzerland.

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in Web and IVR surveys. *Computers in Human Behavior, 19*(1), 1–24.

Van de Kamp, M. E. (2002). Auditory implicit association tests. Unpublished doctoral dissertation. Seattle, WA: University of Washington.

Walton, J. H., & Orlikoff, R. F. (1994). Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research, 37*(4), 738–745.

Wilder, D., & Simon, A. F. (2001). Affect as a cause of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 113–131). Malden, MA: Blackwell.

Yerrapragada, C., & Fisher, P. S. (1993). Voice controlled smart house. In *ICCE 1993: International Conference on Consumer Electronics: Digest of Technical Papers* (pp. 154–155). Piscataway, NJ: IEEE.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2–2), 1–27.

Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology, 43*(2), 347–357.

Zuckerman, M., Miyake, K., & Hodgins, H. S. (1991). Cross-channel effects of vocal and physical attractiveness and their implications for interpersonal perception. *Journal of Personality and Social Psychology, 60*(4), 545–554.